
Linking the Past and the Future

Building a Longitudinal & Predictive Child Support Knowledge Management System

**MAPS Unit
Division of Child Support
Washington State DSHS
P.O. Box 9162
Olympia, WA 98507
(360) 664-5082
FAX (360) 586-3274**

**Semi-Annual Performance Report of the Research Project
Data Warehousing/Data Mining Grant**

**Second Report
February 2003**

**Submitted to the Office of Child Support Enforcement
Administration for Children and Families
U.S. Department of Health and Human Services
Washington, D.C.**

Grant Number 90-FD-0058/01

Linking the Past and the Future

Building a Longitudinal & Predictive Child Support Knowledge Management System

PROJECT SUMMARY ABSTRACT

Our project is a Priority Area I project that uses Data Warehousing and Data Mining and builds capacity of the State agency (Operations and Management of the Program and Financing of the Program). In this conceptual framework, Washington State's Division of Child Support (DCS) is building its capacity to manage and make use of knowledge assets inside the Division to get maximum returns. This approach is especially urgent at a time of shrinking resources and increasing pressure to target dollars as wisely as possible and to perform at an unprecedented level.

To meet the challenge of bringing historical records and multiple databases under one structure, the DCS Management and Audit Program Statistics (MAPS) unit proposed developing a data warehouse. It required acquiring a network-attached storage (NAS), which is the next generation of storage infrastructure. "NAS helps meet changing storage demands by simply attaching to the existing network."¹ NAS boxes are capable of handling heavy database I/O loads. NAS provides an excellent solution for file serving, file sharing and database applications. The NAS is a cost-effective solution to the many problems of large files that are continually expanding as the need for longitudinal data increases, managing volumes of data, hardware failures, viruses, faulty tapes, and lack of scalability for long-term growth.

The second major area of work to be done is to make the data usable. As a recent on-line Fortune article stated, "Businesses are drowning, just drowning in data."² The article talked about the use of data mining to search for patterns and meaning in businesses' data oceans. We were aware that loading data on a NAS would create a better ocean of data, but we still needed to probe the contents of the data for relevant patterns and meaning. MAPS proposed a three-year project not only to bring databases together in one place, but also to bring the collective skills and knowledge of mainframe programmers, researchers, database managers and audit staff together in an unprecedented collaborative effort to build a data warehouse that would have data mining capacity and would improve data reliability.

Progress to Date

An area of emphasis for this grant was to link historic records and to store them for use in decision making, policy analysis, and data analysis on the NAS. There are two distinct efforts currently taking place (*see Figure 1*). First, there is the focus on child support data. The child support data have been brought together from four raw data files: order file, child file, base case file and the child support payment file. Kent Meneghin has taken these four files and created higher summary files, which are analysis ready files

¹ TechRepublic, Dec. 14, 2000, "Gartner Predicts NAS Market to Take Off."

² Stuart F. Brown, Fortune.com, August 2001, "Making Decisions in a Flood of Data."
http://www.fortune.com/indexw.jhtml?channel=artcol.jhtml&doc_id=203525

Linking the Past and the Future

Building a Longitudinal & Predictive Child Support Knowledge Management System

that form the core data warehouse that is used by staff throughout the state. The analysis ready files then feed into the Decision Support System for ad hoc queries by end users and into the Performance Measurement Information System, which provide tables and graphs to end users.

Augmenting this effort are the web enhancements Bryan Enlow has added to the MAPS website that contains the Performance Measures. The Performance Measures were scattered in different places. Bryan was asked to bring everything that has to do with Performance Measures together in one place, allowing easier access to field and headquarters staff and managers.

Most of the measures can be drilled down from the state to the field office level and then to the team and individual worker (Support Enforcement Officer) levels. Another enhancement that has just been added is the ability to select multiple SEOs and field offices for comparison purposes. Future enhancements will allow end users to create graphs by selecting the data they wish depicted by time frame and by level of inquiry (e.g., state, field office, team or Support Enforcement Officer).

Second, there is the bringing together of multiple social service agency data with child support data. Carl Formoso has brought together DCS records from July 1999 to December 2001 to begin the creation of a master identification file (MIF), which will be used to link together longitudinal files created from DCS sources and from external sources. The present file contains both custodial and noncustodial parent identification information including name, social security number, IV-D number, basic individual number, date of birth, and gender. The MAPS unit also has preliminary results on the reliability and utility of each potential identifier.

An example of a longitudinal database that Carl has created includes the linkage of files from 159 State of Washington social service programs. These files are linked to child support records, and have been used for cost avoidance analysis and arrearage analysis to date. The social service data files date from fourth quarter 1993. The most recent social service data obtained were Food Stamp data for 41 months (from January 1999 to May 2002). The Food Stamp data were linked to the January 2001 DCS flat file. (*See the Section on Preliminary Data Mining Work.*)

The infrastructure of equipment and software needed to support the data warehouse/data mining effort is being built and maintained by Ken Forgy and Tom Phan. The infrastructure is comprised of many specialized components that are connected into a communication network that provides storage, application and backup of each workstation and server in the system. (*See schematic of the infrastructure.*)

Data reliability efforts continue. As errors are discovered, the Audit Team works with the Training Unit to go out to the field for retraining staff so that future errors do not occur. The Audit Team has corrected data in the field, conducting training with field staff as the corrections are made so that the knowledge remains in the field.

Linking the Past and the Future

Building a Longitudinal & Predictive Child Support Knowledge Management System

Technical training has been scheduled and taken to enhance the skills of staff, building internal capacity. Other technical training will be taken as it becomes available.

A new process for all Information Technology projects at the Department of Social and Health Services level requires considerable paperwork and levels of approval. The Data Mining project had to complete all the paperwork and be approved, which delayed all purchases until January 2003, about a three-month delay. Primary equipment and software purchases were then sent for acquisition after receiving approval. Additional purchases were submitted the third week of February to ensure we had the equipment and software needed before the state's purchasing freeze begins. The budget climate for the state of Washington continues to become increasingly tense.

Projection of Accomplishments

The next phase is the design of a symmetrical data system that will organize the data into categories: entity, dimension and subject. Krupski Consulting has been given the higher level summary files, file layouts and data definitions, and the html code for the DSS.

Meetings are continuing to further structure the work of Krupski Consulting and the priority of the deliverables. The priorities are to begin with the child support data, then to add external data sources. The first external data source to be added may be the wage history for both noncustodial parents and custodial parents, which dates to 1998 and is refreshed each quarter.

The MAPS unit is discussing whether Krupski Consulting will be using the SAS InterNet interface to create web-based tools for end users to access data and create reports, as well as conduct data analysis. The MAPS unit is also exploring a population matching tool developed by Mr. Krupski that uses biographic information for creating enhanced matches between databases.

In question is whether MAPS will contract with SAS for Pilot Implementation. The contract would speed the successful rollout of the web-based interface. The SAS contractors would work together with Krupski Consulting to create a pilot of one aspect of a front-end application designed for the end user.

Negotiations are still under way to gain access to additional social service data, beginning with Division-level counts of services used by child support custodial parents and noncustodial parents. This access would ease extraction of data. Data sharing agreements are being completed.

Time Line

The project is a three-year project, and will continue through September 29, 2004.

Linking the Past and the Future

Building a Longitudinal & Predictive Child Support Knowledge Management System

Linking the Past and the Future

Building a Longitudinal & Predictive Child Support Knowledge Management System

Linking the Past and the Future

Building a Longitudinal & Predictive Child Support Knowledge Management System

Preliminary Data Mining Work

We have begun to extract information from the data arrays that are presently assembled. We use data from January 1998 to November 2002 monthly DCS extract files and from OFM monthly welfare eligibility records. Quarterly wage records are also available from ESD for this period, but because the reporting of earnings data lags by several quarters, the wage records for 2002 were not considered complete. In addition, we have monthly Medicaid expenses from January 1998 to December 2001 and monthly Food Stamp expenses from January 1999 to May 2002.

We must stress that all results presented here are preliminary – they are not controlled for other factors which could affect the outcomes under discussion. Our planned work will bring in these statistical controls. We do not, however, feel that these statistical controls will change any of the basic ideas presented here, though magnitudes of effects may change.

Custodial Parent Welfare Outcomes

Past research in Washington State has shown that custodial parents classified as receiving regular child support payments subsequently use less welfare and have more employment. This effect occurs after exiting welfare through a lower recidivism rate as well as both a higher rate of finding work and a lower rate of job loss. While on welfare, custodial parents with regular payments appeared identical to custodial parents with irregular payments in terms of rate of welfare exit and in terms of rates of finding and losing work.

The past studies were based on custodial parents who were on welfare during a particular time, and thus do not give a complete view of custodial parents in the DCS caseload. With our data mining efforts, we can now begin to expand on previous work by considering *all* identifiable custodial parents in DCS records from January 1998 forward. The present work involves 417,870 custodial parents.

As in past work, custodial parents are classified with regular payments (CR) when the monthly order amount (MOA) is greater than \$0 and total arrearage debt (TARRS) is less than twice the MOA. Otherwise the custodial parent is classified with irregular payments (CI) or not in DCS records at the time (CN). This classification is made for each of the 59 months of data.

Custodial parents are also classified as on or off welfare in each month. We then create a spell structure in which we determine the number of months (duration) each custodial parent was in a particular state (for example, CR and on welfare) and what event (for example, transition to CR and off welfare) terminated the spell. Table 1 below shows classifications and number of spells observed for each state.

We then bring in earnings data as median quarterly wage during January 1998 to December 2001. This is an indication of earnings capacity, and we define ten levels of capacity as shown in Table 2.

Linking the Past and the Future

Building a Longitudinal & Predictive Child Support Knowledge Management System

Table 1: Cross-Classification and Number of Spells

	Welfare	
	OFF	ON
CR	202,193	29,847
CI	800,111	224,699

Table 2: Earnings Capacity Levels from Median Quarterly Wage

Level	Lower Limit	Upper Limit
1	\$0	\$0
2	> \$0	<= \$1000
3	> \$1000	<= \$2000
4	> \$2000	<= \$3000
5	> \$3000	<= \$4000
6	> \$4000	<= \$5000
7	> \$5000	<= \$6000
8	> \$6000	<= \$7000
9	> \$7000	<= \$8000
10	> \$8000	

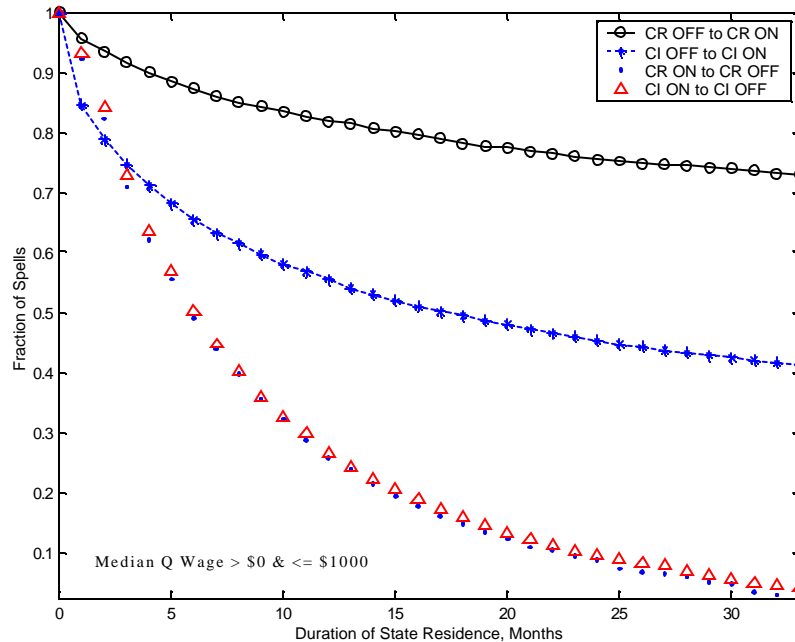
We then compare the durations of spells of CR-Off when terminated by CR-On with durations of spells of CI-Off when terminated by CI-On at each wage level. And we also compare the durations of spells of CR-On when terminated by CR-Off with durations of spells of CI-On when terminated by CI-Off at each wage level.

Figure 2 shows results for level 2 earning capacity where we see that the transition onto welfare is much slower for CR custodial parents, but there is no difference in the durations for welfare exit transitions. The same is essentially true at all wage levels. Even though very little welfare use is seen at higher wage levels the effect of CR appears to reduce rates of welfare entry and has little or no effect on welfare exit rates.

Linking the Past and the Future

Building a Longitudinal & Predictive Child Support Knowledge Management System

Figure 2: Comparing Durations for Welfare Transitions for CR and CI



While welfare use is rare at higher earnings capacity, the effect of CR relative to CI appears almost independent of earning capacity. For each of the 59 months of data, we calculate the fraction on welfare for CR at each wage level (f_{CR}), and the fraction on welfare for CI at each wage level (f_{CI}). To get the relative effect at each wage level we take the difference in these two fractions and then divide by f_{CI} ($[f_{CI} - f_{CR}]/f_{CI}$). At each wage level across 59 months of data CR custodial parents generally use 50 percent to 75 percent less welfare than CI custodial parents.

Custodial Parent Food Stamp Outcomes

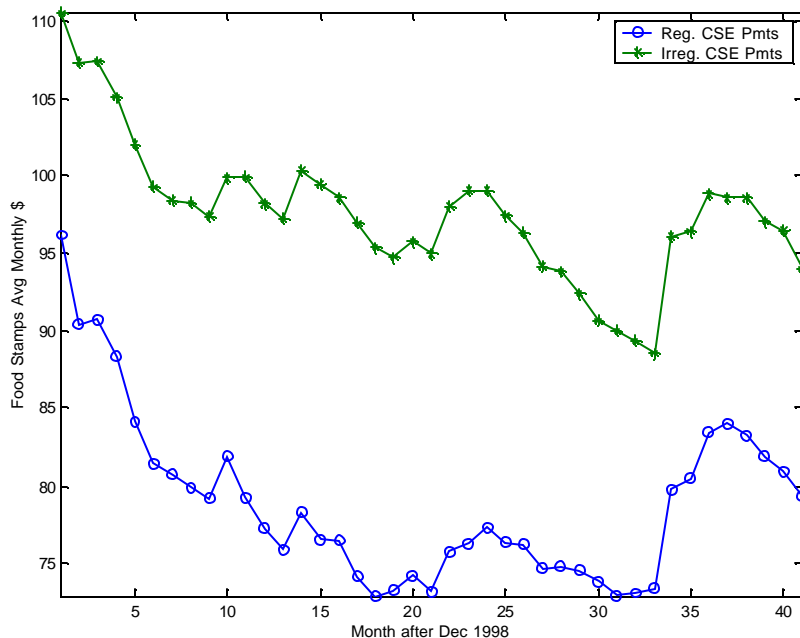
Food Stamp data were linked to the January 2001 DCS extract file. During this period, 105,371 custodial parents from January 2001 show Food Stamp usage.

Figure 3 shows that throughout this 41-month period the average Food Stamps expense for CR (defined as above) custodial parents is less than that for CI custodial parents. The average cost saving approaches 25 percent in May 2000 and June 2000.

Linking the Past and the Future

Building a Longitudinal & Predictive Child Support Knowledge Management System

Figure 3: Comparing Food Stamp Costs for CR and CI



Custodial Parent Medicaid Outcomes

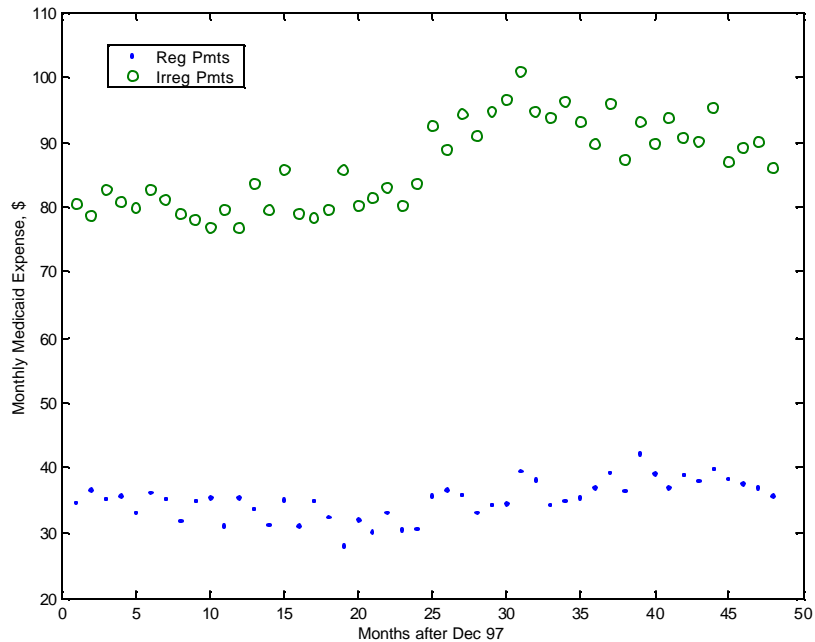
Medicaid cost data were linked to the January 2001 DCS extract file. Preliminary analysis considered all 267,664 custodial parents from January 2001.

Figure 4 shows that throughout this 48-month period, the average Medicaid expense for CR custodial parents (defined as above) is less than that for CI custodial parents.

Linking the Past and the Future

Building a Longitudinal & Predictive Child Support Knowledge Management System

Figure 4: Comparing Medicaid Costs for CR and CI



Noncustodial Parent Outcomes

Washington State has recently completed a study of noncustodial parent arrearage debt. While that study covered a period from October 1993 to June 1997, preliminary analysis with data from January 1998 to November 2002 shows virtually identical results. We will discuss only one aspect of this work here.

There were several indications from the previous work that when MOA was more than about 20 percent of monthly earning arrearages begin to grow. Our results are in line with a recent report from DHHS Office of Inspector General (*'Child Support for Children on TANF,'* OEI-05-99-00392, Feb '02). Table 3 show results with our more recent data which includes 438,112 noncustodial parents.

We see that arrearages grew by \$434 million for those noncustodial parents expected to pay more than 20 percent of their earnings, but arrearage s decreased by \$113 million for noncustodial parents expected to pay 20 percent or less of their earnings. Slightly over 69 percent of all noncustodial parents expected to pay more than 20 percent had a debt increase, while only 27.6 percent of all noncustodial parents expected to pay 20 percent or less had a debt increase.

Linking the Past and the Future

Building a Longitudinal & Predictive Child Support Knowledge Management System

Table 3: Arrearage Outcomes for Noncustodial Parents

Total Arrearage Change millions	all	moa=\$0	moa>\$0
wage>\$0 & mtw>20%	\$434		\$434
mtw<=20%	-\$113	-\$55	-\$58

Percent w Debt Growth	all	moa=\$0	moa>\$0
wage>\$0 & mtw>20%	69.3%		69.3%
mtw<=20%	27.6%	9.6%	38.8%

Arrearage change is the difference between the last recorded arrearage level and the first recorded arrearage level in the 59 months of data. MOA is the average MOA over the 59 months of data. Wage is the average monthly wage from January 1998 to December 2001. MTW is the ratio of MOA to wages.